



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Classifying airports according to their hub dimensions: an application to the US domestic network

Citation for published version:

Hector, R-D, Suau-Sanchez, P & Voltes-Dorta, A 2013, 'Classifying airports according to their hub dimensions: an application to the US domestic network', *Journal of Transport Geography*, vol. 33, pp. 188. <https://doi.org/10.1016/j.jtrangeo.2013.10.011>

Digital Object Identifier (DOI):

[10.1016/j.jtrangeo.2013.10.011](https://doi.org/10.1016/j.jtrangeo.2013.10.011)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Early version, also known as pre-print

Published In:

Journal of Transport Geography

Publisher Rights Statement:

© Hector, R-D., Suau-Sanchez, P., & Voltes-Dorta, A. (2013). Classifying airports according to their hub dimensions: an application to the US domestic network. *Journal of Transport Geography*, 33, 188doi: [10.1016/j.jtrangeo.2013.10.011](https://doi.org/10.1016/j.jtrangeo.2013.10.011)

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Classifying airports according to their hub dimensions: An application to the US domestic network

Héctor RODRÍGUEZ-DÉNIZ¹, Pere SUAUI-SANCHEZ^{2*}, Augusto VOLTES-DORTA³

¹ Héctor RODRIGUEZ-DENIZ

Universidad de Las Palmas de Gran Canaria, FCEE. D.3.01, 35017 Las Palmas de Gran Canaria, Spain.

E-mail: hrodriguez@becarios.ulpgc.es

Tel.: +34 928 458212

² Pere SUAUI-SANCHEZ

Department of Air Transport, Martell House, Cranfield University, Bedfordshire MK43 0TR, United Kingdom

E-mail: p.suausanchez@cranfield.ac.uk

Tel.: +44 1234 75 4227

³ Augusto VOLTES-DORTA

Department of Political Economy, Universitat de Barcelona, Av. Diagonal 690, 08034 Barcelona, Spain

E-mail: avoltes@becarios.ulpgc.es

* Corresponding author: Pere SUAUI-SANCHEZ

Department of Air Transport, Martell House, Cranfield University, Bedfordshire MK43 0TR, United Kingdom

E-mail: p.suausanchez@cranfield.ac.uk

Classifying airports according to their hub dimensions: An application to the US domestic network

Abstract

Government agencies classify airports for different purposes, including the allocation of public funding for capacity developments. In a context of hub classification, determining the contribution of each airport to the national network in terms of the two dimensions of “hubbing”, i.e. traffic generation and connectivity, is a key aspect. However, the choice of an appropriate indicator of airport connectivity is still an unresolved issue. This paper contributes to the existing literature by adapting the well-known flow centrality indicator to an air transport context and thus developing a novel measure of airport connectivity. An application to the US airport network is provided, using quarterly data on passenger demand to perform a detailed time-series analysis of airport connectivity patterns between 1993 and 2012. The suitability of our flow-based indicator, against other commonly used centrality measures, is assessed by testing their sensibility to the major cases of airline de-hubbing in the US. The flow-based indicator is then used to define an alternative airport classification method within the context of the Federal Aviation Administration’s National Plan of Integrated Airport Systems (NPIAS). Results show that there is potential for improving the existing regulatory airport classification by taking connectivity into consideration.

Keywords

Airport networks, connectivity, flow centrality, hierarchical clustering.

Classifying airports according to their hub dimensions: An application to the US domestic network

1. Introduction

The US Federal Aviation Administration (FAA) estimates that \$42.5 billion will be available over the period 2013-2017 to fund infrastructure developments for all segments of civil aviation under the Airport Improvement Program (AIP). The National Plan of Integrated Airport Systems (NPIAS) is used by the FAA in administering the AIP. In the NPIAS (FAA, 2011), investment requirements and funding priorities are set according to an airport typology based on each airport's traffic share over total US passenger enplanements (Table 1). By using enplanements instead of passengers, the FAA is trying to consider the importance of transfer traffic, but, while the merit and simplicity of that approach is not questioned, it does not fully respond to the rise and dominance of hub-and-spoke networks. Hub-and-spoke operations are typically achieved by consolidating originating and transfer passenger flows (Doganis, 2010; Button 2002), which implies the existence of two dimensions of "hubbing": traffic generation and connectivity. Since one of the main objectives of the AIP is to fund airport capacity expansions in order to reduce congestion and delays, from a social perspective, it seems reasonable that funding priority should be given to airports playing a central role in the network, not just because they have significant proportion of US traffic, but also because passengers are connecting through them to other destinations. Hence, there is a potential for the FAA, as a public agency, to optimize the social benefits from AIP investments by improving the NPIAS airport classification method to acknowledge the importance of hub connectivity along with the airports' potential for traffic generation.

This paper tries to contribute to the debate on airport classification for policy purposes by focusing on the US, with special regard to the National Plan of Integrated Airport Systems. We bring together the dimensions of traffic generation and connectivity and

develop two flow-based indicators that are used to define an alternative airport classification method, which could have potential interest for the FAA.

The paper is structured as follows: Section 2 provides reviews airport classifications, airport connectivity and centrality indicators. Section 3 covers the data and methodological aspects, including the development of a flow-based centrality indicator to measure airport connectivity. Finally, Section 4 presents the results of the benchmarking exercise of centrality indicators in order to test the suitability of our flow-based connectivity indicator, also the benefits of classifying large airports according to their hub dimensions are discussed and an alternative classification of large US hubs is provided using hierarchical clustering techniques.

Table 1. Commercial airport categories according to FAA's current classification. Source: FAA.

Commercial Airport Type At least 2,500 boardings	Hub type Percentage of annual passenger boardings	Common name
Primary	Large 1% or more	Large Hub
	Medium At least 0.25%, but less than 1%	Medium Hub
	Small At least 0.05%, but less than 0.25%	Small Hub
	Nonhub More than 10,000, but less than 0.05%	Nonhub Primary
Nonprimary	Nonhub At least 2,500 and no more than 10,000	Nonprimary Commercial Service

2. Airport classification, hub dimensions and connectivity

2.1 Airport classification

Airport classification into homogeneous groups is typically used for benchmarking purposes in both policy and management contexts. It is a good starting point for the analysis of a variety of issues, such as the impact of air route deregulation, airport congestion, suitable development policies and regulatory norms and airport performance analysis (Malighetti et al., 2009). Then, the type of variables (e.g. Jessop, 2012) and the specific technique to

classify airports heavily depends on the objective of the study. Previous literature on airport classification is very heterogeneous, although it seems to be a consensus that hierarchical clustering methods are the most commonly employed (Rodríguez-Déniz and Voltes-Dorta, 2014). These have been applied to a wide variety of subjects, ranging from accessibility and connectivity (Burghouwt and Hakfoort, 2001; Malighetti et al., 2009), runway geometry (Galle et al., 2010), slot allocation (Madas and Zografos, 2008), and the comparative analysis of efficiency and productivity (Sarkis and Talluri, 2004).

With regard to the US, the closest reference to the present paper is the contribution by Adikariwattage et al. (2012). They classified airports in the US domestic network using four variables: number of boarding gates, number of origin and destination passengers, transfer and international passengers. They cluster airports in two steps, separating the number of gates from the passenger volumes leading to nine groups that combine all these variables. While the merit of that classification method is not challenged here, their results are not particularly sensitive for the largest hubs, since all of them are grouped together the same category (e.g., JFK, LAX, ATL, and CLT), despite presenting radical differences in their hub profiles as it is analysed in Section 4. We build on the contribution by Adikariwattage et al. (2012) to produce a more sensitive method for classifying and discriminating large hubs within the context of the NPIAS. We try to achieve this by focusing on the airports' contribution to the network in terms of both traffic generation and connectivity, rather than simply relying on passenger volumes.

2.2 Hub dimensions, connectivity and airport classifications

Hub-and-spoke operations are typically achieved by consolidating originating and transfer passenger flows (Doganis, 2010; Button 2002), which implies the existence of two dimensions of “hubbing”: traffic generation and connectivity.

Connecting traffic is traffic between airport A and airport B via the hub airport H. Effective hubbing then generates substantial volumes of additional traffic at the hub airport. The city-pair coverage that can be obtained is significant, since increase in the number of airports served from the hub impacts exponentially on the number of city-pairs served (Doganis, 2002).

Generated traffic is traffic between hub airport H and airport A. Although we tend to focus on the importance of transfer traffic at hubs, these are still highly dependent on non-transfer traffic, since some flight sectors have important shares of non-transfer passengers and the increase of direct services at the hub can produce a multiplying effect on the generation of traffic from and to the hub. As a matter of fact, most hubs are located in regions with large local markets (Liu et al. 2006).

Concerning specifically airport hub classification and identification, it is difficult to find studies using both dimensions of airport hubbing (i.e., traffic generation and connectivity).

Some connectivity measures¹ are able to capture, to some extent, both the generation of traffic and the transfer traffic. Yet, since they rely on supply data of seats and frequencies, connectivity indices usually focus on different aspects of potential connectivity, such as the number of feasible connections or transfer opportunities available to the passenger, and centrality indices evaluate the airport's hubbing potential on the basis of its central location in the network. This is related to the difficulties in collecting demand data on actual connections

¹ See Burghouwt and Redondi (2013) for an extensive review of this type measures. These indicators can be roughly classified according to whether they consider temporal restrictions (to determine when an indirect connection is viable) or take into account all possible connections in the network (global versus local models). While the bulk of the literature is focused on time-dependent local measures (e.g., Doganis and Dennis, 1989; Dennis, 1994a, 1994b; Bootsma, 1997; Veldhuis, 1997; Danesi, 2006; Burghouwt, 2007; Budde et al., 2008; Matsumoto et al., 2008; Suau-Sanchez and Burghouwt, 2012), there has been a growing interest on global models in the recent years (e.g., Guimerà et al., 2005; Guida and Maria, 2007; Bagler, 2008; Cronrath et al., 2008; Malighetti et al., 2008; Reggiani et al., 2008; Xu and Harris, 2008; Paleari et al., 2010; Berger et al., 2011; Wang et al., 2011; Zeng et al., 2011; Jia and Jiang, 2012). Global models are usually based on measures coming from complex network theory (e.g., Freeman, 1977, 1978), which implementation is more demanding on both data and computation.

made by the passengers, which would make the analysis much more relevant from a regulatory perspective (e.g., quantifying how many passengers benefit from AIP investments).

In this regard, Adikariwattage et al. (2012) consider the two-hubbing dimensions using actual demand data, but they do not integrate them in a single demand-based index. Hence, to our knowledge there is no demand-based measure of connectivity beyond the number and proportion of connection passengers.

The necessary information on actual passenger routings, however, has been made available for the US domestic network by the Department of Transportation (Airline Origin and Destination Survey DB1B, RITA (2013)). This database includes a 10% of tickets sold; hence, it does not allow us to measure the total numbers of originating and connecting passengers at each airport, *a priori* the obvious indicators for traffic generation and connectivity. Alternatively, we aim to develop two indicators (traffic generation and connectivity) that are based on actual, instead of potential, hubbing activity. These indicators are eventually merged into a single demand-based index.

Hence, by conceptualizing passenger trips as “flows”, this paper tries to contribute by adapting a well-known flow centrality indicator to an air transport context and develop a novel measure of airport connectivity. The suitability of our flow-based connectivity indicator is assessed against other commonly used centrality measures by testing their sensibility to the major cases of airline de-hubbing in the US. In order to carry out this benchmarking exercise, we use quarterly data on passenger demand to perform a demand-based time-series analysis of airport connectivity patterns between 1993 and 2012. Based on our flow-based indicator, we define an alternative airport classification method with stronger hub discrimination power than the existing FAA airport classification.

3. DATA AND METHODOLOGY

3.1 Database

As mentioned above, we use of the publicly available data provided by the Bureau of Transportation Statistics of the Research and Innovative Technology Administration (US Department of Transportation). The Airline Origin and Destination Survey (Database code: DB1B) (RITA, 2013) is a sample of airline ticket information from more than 30 US carriers. The survey covers about 10% of domestic tickets sold by the reporting carriers with specific indication of the full itinerary for multi-sector journeys. Additional variables included in the dataset are the operating carrier, the number of passengers or the distance flown, among others. These records are available on a quarterly basis and were collected from the first quarter 1993 to the second quarter 2012 for our time-series analysis. The resulting sample contains about 350 million records representing individual itineraries.

It is worth clarifying that only domestic itineraries are included in this database (i.e., journeys with both origin and destination airports located in the US) and that there are not available databases providing information on the full itinerary of international passengers. Therefore, for the interest of data consistency we have worked only with data of the DB1B database.

3.2 Flow centrality

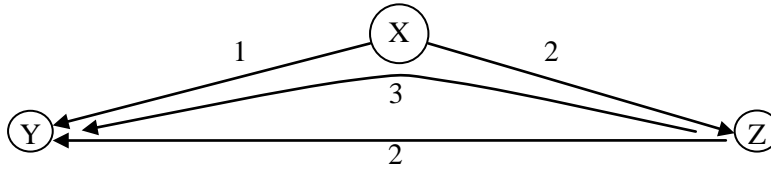
In order to measure airport connectivity, this paper adapts the well-known flow centrality measure from Freeman et al. (1991). This indicator was developed in a social network context and aims to quantify the proportion of the maximum directed flow of information (m) between two nodes (j,k) that travels through an intermediate node (x_i). This maximum flow will depend on the capacity of the links in the network and it is calculated for each pair of nodes by applying some simple rules, such as that incoming flow must equal outgoing flow

for all nodes involved in the transmission of information. By aggregating all possible pairs of nodes (j,k) , the measurement of flow centrality for node x_i is easily calculated as the total directed flow that passes through x_i divided by the total flow between all pairs of nodes where x_i is neither a source of information nor its final destination. Thus, the flow centrality (valued between 0 and 1) measures the proportion of the total network flow that travels through x_i .

$$(1) \quad C'_F(x_i) = \frac{\sum_{j < k}^n \sum_k^n m_{jk}(x_i)}{\sum_{j < k}^n \sum_k^n m_{jk}}$$

Adapting this indicator to an air transport context is straightforward. Airports in the US domestic network are defined as nodes. The links that connect the nodes are the individual flight sectors operated by airlines. Passenger traffic is the flow that travels through the network between a point of origin (j) and a final destination (k) using a variety of routes (either non-stop or connecting). Note the market-based definition of passenger flow. The capacity of the links is defined by the total passengers from all different origin/destination markets that share the same individual sector. Since the available data provides information on origin, destination, and intermediate airports (when applicable) at a passenger level, it is possible to obtain both flow and capacity matrices. By incorporating all these definitions into the C'_F formula (1) and assuming that the maximum flow equals observed flow, the degree of flow centrality for airport x_i collapses into a quotient between total number of passengers that connect through x_i and total network passengers that travel in all markets that do not start or terminate at x_i . This ratio becomes our flow-based measure of connectivity. A numerical example is provided in Figure 1, where numbers denote passengers in each market meaning that the market between Y and Z airports comprises 5 passengers, 2 travelling non-stop and 3 via the hub X . Therefore, the value of flow centrality for airport X is $3/5$ (0.6). In other words, the network has a 60% dependence on X to serve their markets.

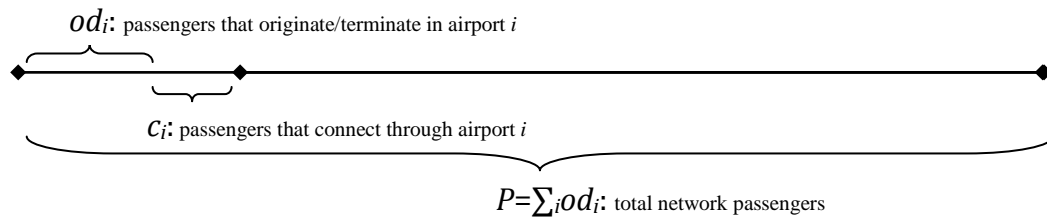
Figure 1. Numerical example of flow centrality



3.3 An aggregated indicator for the hub dimensions

The flow-based centrality indicator will be used to develop an alternative airport typology. This is expected to be most useful to classify large airports with a potential to serve connecting traffic flows. However, it is worth remembering that connectivity is only one of the two main dimensions of a hub, which should also generate a significant amount of traffic (either as origin or final destination) that allows the airlines to consolidate services and exploit economies of density (Caves, 1997). These two dimensions of hub airports (connectivity and traffic generation) will become the variables of our proposed classification method. In this regard, Figure 2 shows how network flows can be partitioned for each airport.

Figure 2. Partition of the total network flows with respect to the i -th airport.



Following the simple nomenclature presented in Figure 2, we can easily define two separate measures for each airport's traffic contribution to the network. The first one (OD_i) is calculated as the ratio between the passengers that originate or terminate at the i -th airport (od_i) and the total network passengers (P). This serves as an indicator of the airport's importance as generator of traffic. The second measure is the flow-based indicator (renamed

C_i) that measures the airport's importance as a connecting point. As defined in the previous section, it is calculated as the ratio between connecting passengers (c_i) and total network passengers that do not originate or terminate the i -th airport ($P - od_i$).

$$(2) \quad OD_i = \frac{od_i}{P} \quad C_i = \frac{c_i}{P - od_i}$$

These two indicators can be used to obtain a more detailed profile on the individual airports' hub characteristics that can be used to develop a typology of airports in the US. Furthermore, it is also possible to establish a link between these measures and the aggregated indicator currently used by the FAA. Since the FAA considers enplanements instead of passengers for their indicator, we just need to define the total number of enplanements in the network (E) and the sum of all types of traffic ($od_i + c_i$) across all the airport population. Note the multiple-counting of connecting passengers (which implies that $E > P$). Then, the FAA indicator (FAA_i) is defined as the i -th airport's traffic share over total enplanements.

$$(3) \quad E = \sum_i (od_i + c_i) \quad FAA_i = \frac{od_i + c_i}{E}$$

Therefore, we can establish the following relationship between the FAA indicator and the disaggregated ones:

$$(4) \quad FAA_i = OD_i \frac{P}{E} + C_i \frac{(P - OD_i)}{E}$$

Equation 4 will be used in Section 4.2 in order to map the different combinations of OD_i and C_i that lead to the same value for FAA_i . This is expected to show the pitfalls of the aggregated system for hub classification.

3.4 Hierarchical clustering

Our alternative classification criteria will be expressed as a set of threshold values for connectivity and traffic generation, determined by using agglomerative hierarchical

clustering (AHC)² on a cross-section of our airport sample for the year 2011. The existing literature indicates that AHC has been the most popular choice to classify airports, yet a great degree of ad-hoc procedures are still used (Rodríguez-Déniz and Voltes-Dorta, 2014). The resulting hierarchical classification is typically presented in a tree-like diagram (i.e. dendrogram) that provides a much more informative structure than the flat clusters obtained from other partitioning methods, such as k -means. Starting from a matrix of pair-wise distances between the individual objects, AHC performs a sequence of merge operations that produce additional clusters at new levels of aggregation and are governed by a predefined clustering strategy. This paper uses the complete-linkage algorithm, combined with a Euclidean distance metric. In this method, each step merges the nearest two clusters according to the farthest distance among their components, which leads to more compact aggrupations. Hierarchical methods do not require predefining the number of clusters, which can instead be identified by using a “tree-cutting” method. We employ the pseudo- F coefficient that takes the ratio of between-cluster variance to within-cluster variance (Calinski and Harabasz, 1974). The edges of the resulting clusters are then used to define the thresholds of our new airport categories.

4. Results and discussion

4.1 Flow centrality sensibility analysis results

Before proceeding with the analysis, we prove the validity of our demand-based flow centrality measure by testing its sensibility to changes in airport connectivity. One of the most sudden changes in airport connectivity are de-hubbing cases, when a dominating carrier dismantles its hubs activities in one of its main bases (Bhadra, 2009). Doing a supply-based time-series analysis, Redondi et al. (2012) identify up to 37 worldwide cases of de-hubbing

² General references to data clustering are Everitt et al. (2001) and Xu and Wunsch (2005).

from 1997 to 2009. Using their list, we apply four different centrality indicators (Degree Centrality [Degree], Weighted Betweenness Centrality [WBC], Un-weighted Betweenness Centrality [BC], and Flow Centrality [Flow-C_i]) for a selection of US airports that have suffered a de-hubbing process during the last decades. For this sensitivity analysis time-series data was adjusted for seasonality.

Degree centrality (Nieminem, 1974) represents the number of connections that an airport has. It has become a standard approach for measuring the connectivity potential of every node in the network, being strongly correlated to the airport's passenger throughput. Degree centrality can be formalized for an airport i as:

$$(5) \quad C_D(i) = \sum_j \frac{a_{ij} + a_{ji}}{2}$$

where a_{ij} is the adjacency matrix, in which $a_{ij} = 1$ if the airport i is connected to airport j , and 0 otherwise. Betweenness centrality (Freeman, 1977) quantifies the prominence of an actor in terms of connectivity within a network by computing how frequently a node lies on the shortest path between any other two nodes. The betweenness centrality measure is given by:

$$(6) \quad C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where σ_{st} is the number of minimum length paths connecting nodes $s \in V$ and $t \in V$, and $\sigma_{st}(v)$ is the number of such paths in which some $v \in V$ lies on. Airports with high levels of betweenness are strategically placed close to the major airline markets and therefore they will be in a privileged, central position in comparison with the rest of their peers. From an air transport perspective, however, the betweenness centrality presents some serious drawbacks due to its strong topological motivation. In order to overcome these limitations, Rodríguez-Déniz (2012) introduced a market-based betweenness centrality to identify key airports in an air transport network according to both their topological position (i.e. connectivity potential) and the relevance of the markets they serve in terms of traffic density, defined as:

$$(7) \quad C_{B_{mkt}}(v) = \sum_{s \neq v \neq t \in V} \frac{Q_{st}}{Q} \cdot \frac{\sigma_{st}(v)}{\sigma_{st}},$$

where (Q_{st}) is the total number of passengers that travelled on market $s, t \in V$, and (Q) the total number of passengers in the sample. As a result, top ranked airports are likely to play an important role within the network by combining a central location with relevant market service. Airports lacking of either characteristic will be probably mid-ranked. Airports with similar traffic levels will be classified according to their centrality.

Table 2. Percentage loss of centrality for a selection of de-hubbing cases.

<i>Start Year & Quarter</i>	<i>End Year & Quarter</i>	<i>Airport</i>	<i>Hub carrier</i>	<i>Main cause</i>	<i>Degree</i>	<i>BC</i>	<i>WBC</i>	<i>Flow-C_i</i>
2005 Q4	2010 Q4	Cincinnati (CVG)	Delta-Northwest	Merger	-16.62%	-43.89%	-35.92%	-80.38%
2005 Q2	2005 Q4	New Orleans (MSY)	-	Hurricane Katrina	-19.36%	-17.99%	-40.66%	-82.37%
2001 Q4	2005 1	Pittsburgh (PIT)	US Airways	Network Restructuring	-5.89%	13.93%	-4.57%	-77.91%
2001 Q3	2004 Q1	Saint Louis (STL)	American-TWA	Merger	-7.26%	4.72%	-9.22%	-83.50%
2001 Q3	2001 Q4	Reagan (DCA)	US Airways	9/11 Security Restrictions	-6.99%	29.30%	-11.84%	-73.91%
2001 Q2	2001 Q4	Raleigh-Durham (RDU)	Midway	Bankruptcy	-8.80%	38.56%	-21.90%	-81.55%
1997 Q1	1997 Q4	Colorado Springs (COS)	Western Pacific	Network Restructuring	-5.29%	-1.78%	10.13%	-77.74%
1995 Q1	1996 Q1	Nashville (BNA)	American	Network Restructuring	-2.89%	25.11%	0.90%	-72.15%

Degree: Degree Centrality.

BC: Un-weighted Betweenness Centrality.

WBC: Weighted Betweenness Centrality.

Flow-C_i: Flow Centrality.

Table 2 shows the results, which vary widely across the four indicators, illustrating the numerous ways in which centrality is measured and the impact of these conceptual differences on their characterization of airport connectivity. Unsurprisingly, degree centrality, which depends solely on the airport's number of connections without taking into account route density, is the indicator that shows the least variability. This is explained by the practice of de-hubbed carriers and alliances to keep a minimum service in order to prevent re-hubbing

by rival alliances (Redondi et al., 2012). Weighted and un-weighted betweenness centrality are also highly dependent on the airports' geographical location and route structure (see Wang et al. (2011) who get a similar effect for China), although results are much more erratic and unpredictable. While airports such as Cincinnati and Washington Reagan show the expected drop of centrality linked to the closure of direct air routes, it is difficult to explain why Pittsburgh, Colorado Springs or Nashville experienced a significant increase in betweenness centrality during their de-hubbing period. Contrary to the other indicators, flow-based centrality is the only indicator that clearly presents the expected negative signs in all cases.

In addition to Table 2, the lack of sensibility of degree and betweenness indicators to airline de-hubbing is shown graphically in Figure 3, which shows the normalized results for the massively de-hubbed St Louis Airport (STL) over the whole sample period. Aside from seasonal variations, both degree and un-weighted betweenness centrality do not appear to change significantly. This is consistent with the topological nature of both indicators, which, as we mentioned above, are heavily dependent on the airport's fixed location and route structure. Only when market-based weights are applied to the topological indicators it is possible to see a slight long-term decrease in centrality (WBC).

Figure 3. Evolution of centrality measures at St Louis International Airport (STL), 1993-2012.

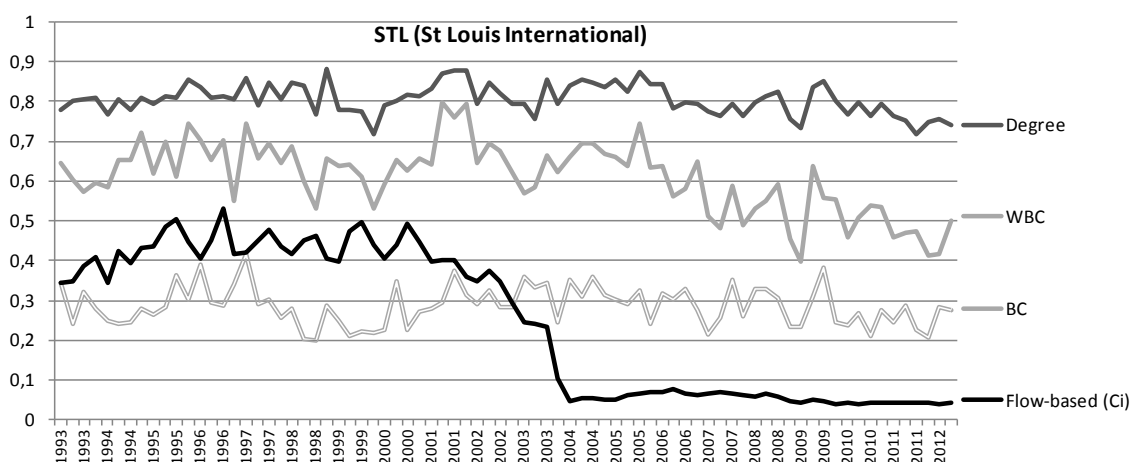


Figure 4 shows the evolution of flow centrality at other US airports with de-hubbing events, being clear that the flow-based centrality indicator behaves in a similar way than in Figure 3. Furthermore, Figure 4 shows that the value of flow centrality as measure is not only limited to big changes in the in the network structure, but it also reacts well to punctual events, such as industrial actions, in which the flow of traffic is interrupted. It is also important to highlight that de-hubbed airports do not tend to recover after the airline has completed the process, thus agreeing with the supply-side analysis by Redondi et al. (2012).

Figure 4. Evolution of flow-based centrality at selected airports 1993-2012.

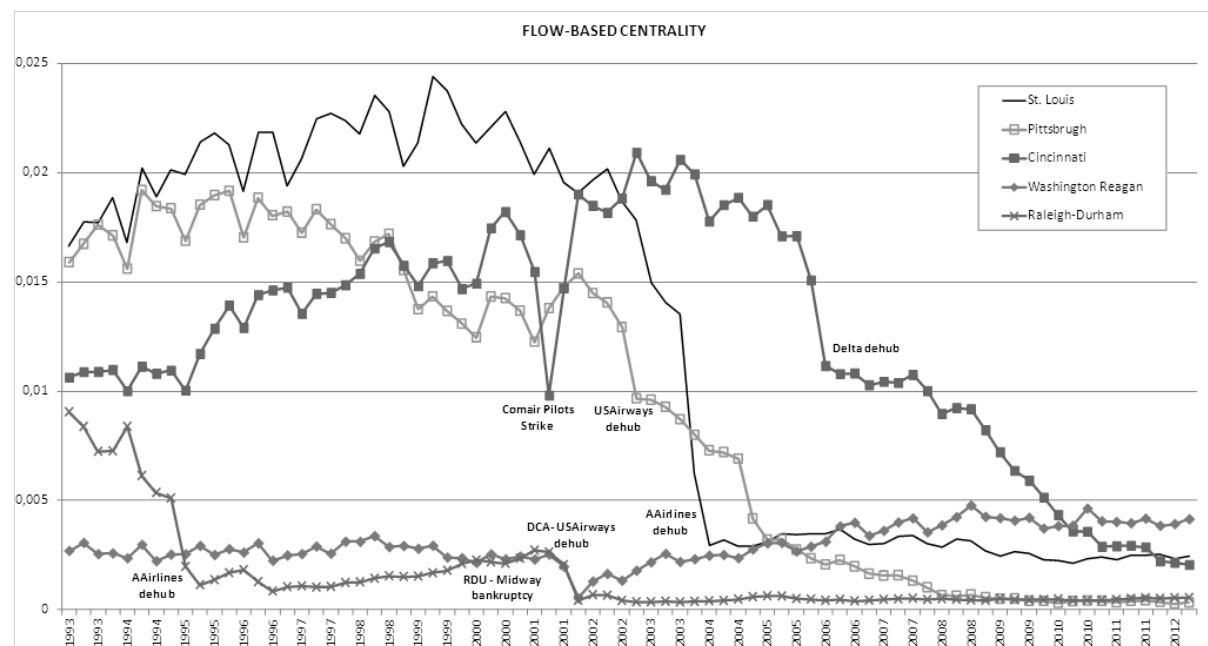


Figure 5. Evolution of flow-based centrality at large airports 1993-2012.

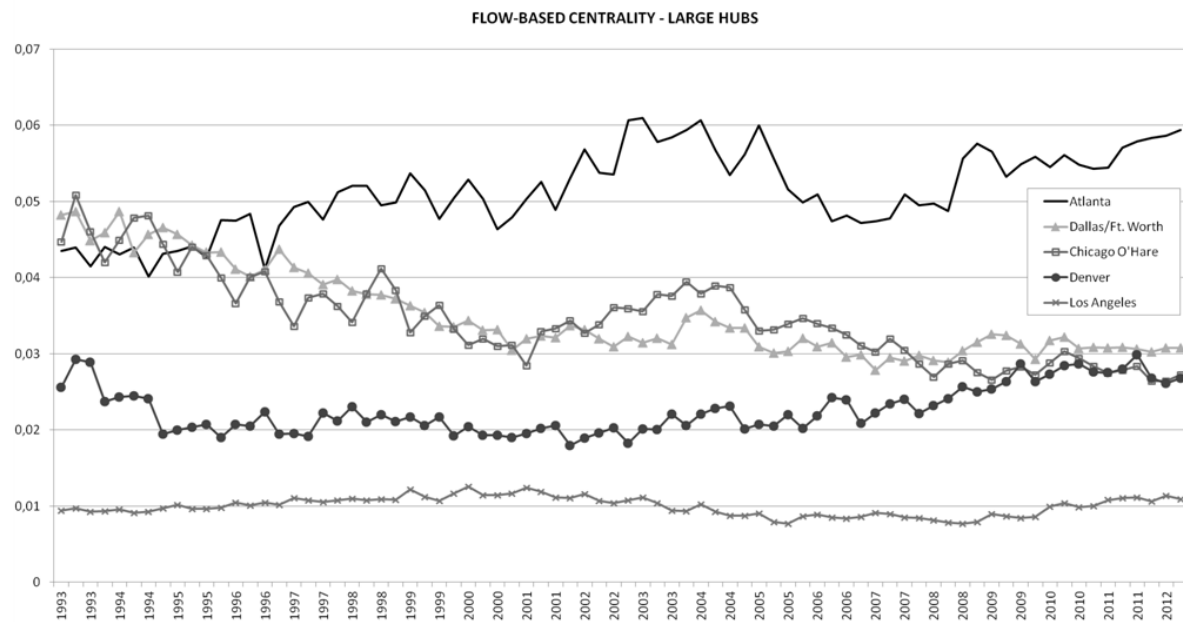


Figure 5 provides the same time-series analysis for the largest airports in the NPIAS and helps showing the capacity of the indicator to discriminate between airports in the US network. Atlanta clearly stands out as a first-class connecting gateway right after the 1996 Summer Olympics and it has steadily increased its relevance ever since, most recently scoring a 6% flow-based centrality. This means that Atlanta is serving around 6% of total US passengers that originate elsewhere, thus indicating a massive dependence of the US domestic network on this particular hub. On the contrary, note the evolution of Dallas/Ft. Worth and Chicago O'Hare, whose relevance, in terms of centrality, has steadily decreased over the last decades, joining Denver in a hypothetical 2nd tier. Finally, the discriminatory power of flow-based centrality is seen for Los Angeles, that has steadily kept a relatively low connectivity level of around 1%. This is consistent with the fact that Los Angeles is the world's busiest airport for originating traffic, but it is not a particularly strong "connecting" hub, at least in relation to the other airports in Figure 5. Thus, the second indicator for traffic

generation (OD_i) reveals to be necessary to fully characterize the different roles played by the main commercial airports in the US.

Hence, we can conclude that the direct relationship between the changes in the amount of connecting traffic and the changes in the flow centrality measure results shows that this indicator is a sensitive measure of airport connectivity.

4.2 Classifying airports according to their hub dimensions: an application to the NPIAS

Having tested the sensitivity of $Flow-C_i$, we can then proceed to calculate the generation of traffic indicator (OD_i), the flow-based indicator (C_i) and the aggregated FAA indicator (FAA_i) for the whole sample. Table 3 and figures 6 and 7 present the results on the two hub dimensions and the aggregated FAA_i for all FAA-designated large (1% or more) and medium hubs (between 0.25% and 1%). Using Equation 4, we are also able to represent the different levels of the FAA indicator as a combination of connectivity and traffic generation. This graphical representation allows for a better comparison between both classification dimensions.

At first sight, we can conclude that the definition of a 1% share of enplanements as a threshold for large hubs is appropriate since it is located around a natural breaking point in the dataset. This is undoubtedly a first advantage of the FAA classification, and the second one is, evidently, its simplicity, as it only depends on a simple ratio. However, simplicity comes at the cost of discriminating power. All airports above 1% are large hubs, but major differences in terms of generation and connectivity exist among them (Figure 6). For example, in the same category, the current FAA system mixes a mid-size hub (Charlotte Douglas-CLT) with a massive one (Atlanta-ATL), whose contribution to the network is twice as large in both dimensions, and both of them are joined by a massive traffic generator (Los

Angeles-LAX). Thus, when aggregating both hub dimensions into a single indicator, the current FAA airport classification (Table 1) cannot discriminate among the different airports.

Table 3. Traffic generation and connectivity hub dimensions, and aggregated FAA indicator for medium and large hubs, 2011.

	<i>ODi</i> (%)	<i>Ci</i> (%)	<i>FAAi</i> (%)		<i>ODi</i> (%)	<i>Ci</i> (%)	<i>FAAi</i> (%)		<i>ODi</i> (%)	<i>Ci</i> (%)	<i>FAAi</i> (%)		<i>ODi</i> (%)	<i>Ci</i> (%)	<i>FAAi</i> (%)
ATL	5.60	5.70	4.64	FLL	4.00	0.09	1.73	RDU	1.76	0.05	0.77	ONT	1.02	0.02	0.44
ORD	5.91	2.76	3.60	EWB	3.59	0.37	1.67	SJC	1.75	0.06	0.76	OGG	0.93	0.09	0.43
DEN	5.78	2.81	3.57	SAN	3.46	0.12	1.51	MSY	1.75	0.05	0.76	BUR	1.00	0.02	0.43
LAX	7.29	1.09	3.51	DCA	3.18	0.40	1.51	MKE	1.51	0.20	0.72	PVD	0.87	0.01	0.37
DFW	4.78	3.06	3.26	MD	2.60	0.84	1.45	SAT	1.64	0.04	0.71	OM	0.85	0.02	0.36
LAS	6.77	0.70	3.14	TPA	3.20	0.15	1.41	PIT	1.60	0.04	0.69	RNO	0.76	0.03	0.34
PHX	4.84	1.90	2.82	SLC	2.19	0.99	1.34	RSW	1.60	0.01	0.68	TUS	0.76	0.02	0.33
MC	6.20	0.24	2.72	PDX	2.40	0.21	1.10	DAL	1.27	0.29	0.66	ANC	0.66	0.08	0.31
SFO	5.43	0.63	2.55	HNL	2.27	0.28	1.08	IND	1.49	0.03	0.65	OKC	0.72	0.02	0.31
SEA	4.66	0.75	2.28	IAD	1.91	0.61	1.06	CLE	1.21	0.29	0.63	ORF	0.67	0.01	0.29
BOS	4.98	0.09	2.14	MIA	2.14	0.29	1.03	SJU	1.33	0.03	0.57	SDF	0.65	0.02	0.28
CLT	2.04	2.98	2.10	STL	2.17	0.25	1.02	CMH	1.27	0.03	0.55	RIC	0.65	0.01	0.28
LGA	4.67	0.20	2.05	MCI	1.96	0.12	0.88	ME	0.70	0.59	0.55	LGB	0.64	0.02	0.28
MSP	3.40	1.46	2.04	OAK	1.91	0.10	0.85	PBI	1.24	0.02	0.53	GEG	0.64	0.01	0.27
PHL	3.53	1.10	1.94	HOU	1.65	0.31	0.83	BDL	1.19	0.01	0.51	MHT	0.62	0.00	0.26
DT	3.08	1.44	1.90	SNA	1.91	0.04	0.82	JAX	1.14	0.03	0.50	ELP	0.59	0.03	0.26
JFK	3.95	0.31	1.80	SMF	1.87	0.06	0.82	ABQ	1.08	0.09	0.49	BHM	0.58	0.03	0.26
BWI	3.60	0.67	1.80	AUS	1.82	0.06	0.80	CVG	0.85	0.27	0.47	BOI	0.58	0.02	0.25
IAH	2.65	1.60	1.78	BNA	1.68	0.20	0.79	BUF	1.09	0.02	0.47	TUL	0.56	0.02	0.25

Figure 6. Disaggregated vs. FAA airport classification: large hubs (>1%), 2011.

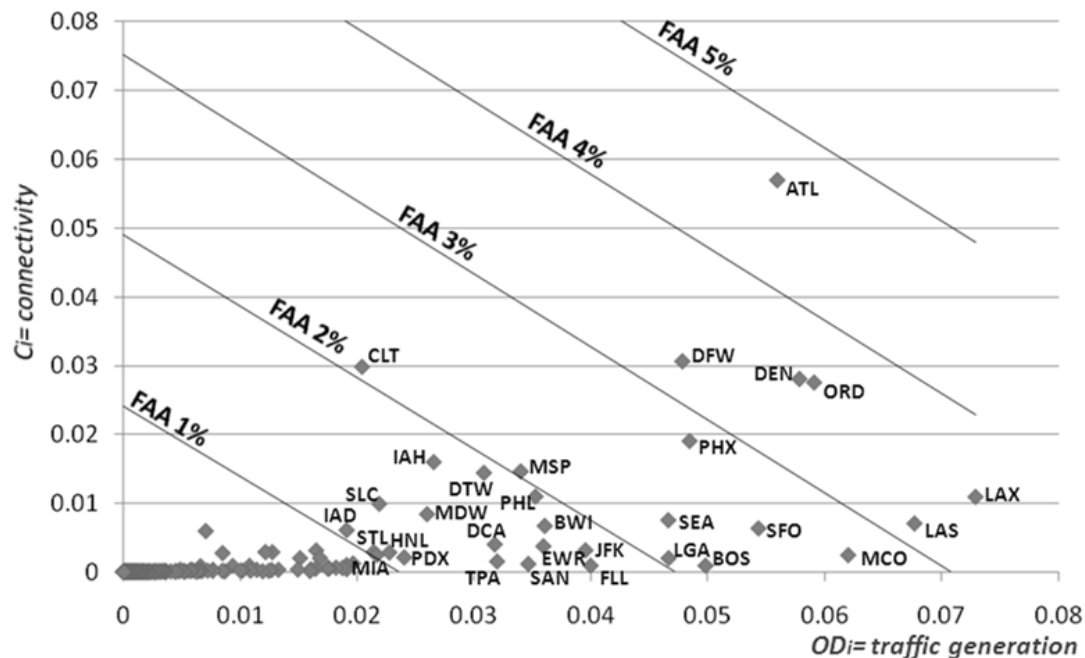
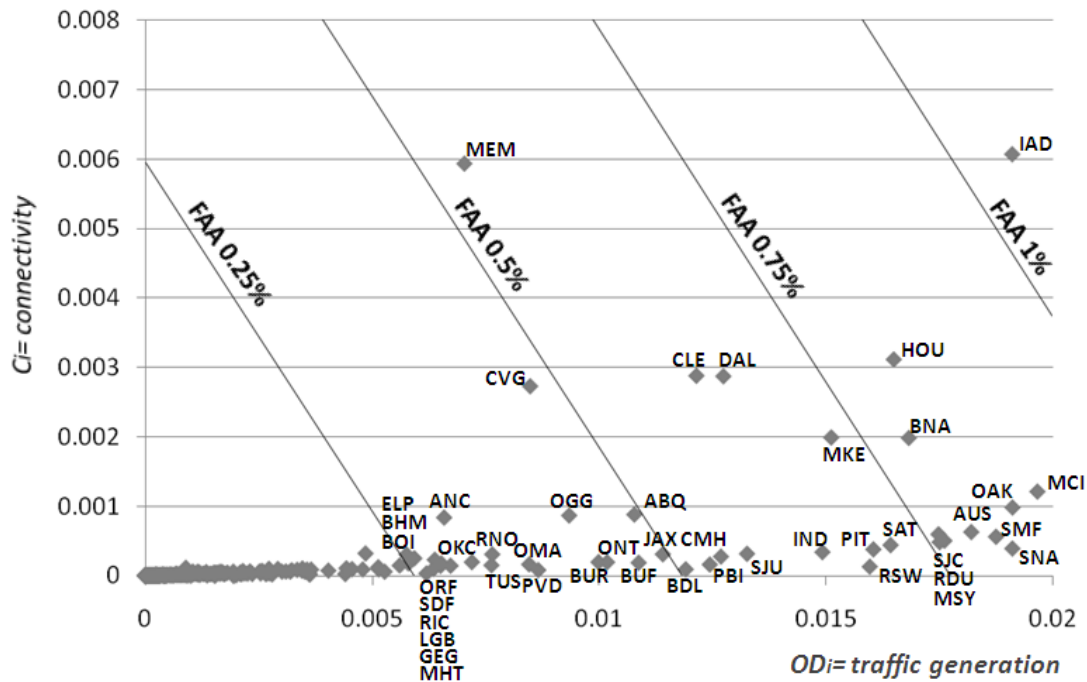


Figure 7. Disaggregated vs. FAA airport classification: medium hubs (0.25-1%), 2011.



In order to obtain an alternative airport classification we use the agglomerative hierarchical clustering on the basis of the generation of traffic and the flow-based indicators. The results of the clustering are presented in Table 4 and 5 and the full dendrogram is provided in Appendix A. The optimal truncation level (similarity=0.0182) leads to nine clusters. However, for simplicity, we decided to explore the dendrogram for the immediately next level of aggregation (0.03), leading to six clusters and a much easier interpretation of results (Figure 8 and Table 6). Following the previous example, now, with this alternative airport classification, Atlanta Airport and Charlotte Douglas Airport would be placed in their own categories –first and third tier hubs respectively–, which is not surprising since there are no other airports that get close to their hub profiles. The remaining airports that score high in both dimensions, such as Dallas/Fort Worth (DFW) or Chicago O’Hare (ORD) are classified as second tier hubs. Criteria for belonging to these clusters are detailed in Table 6.

Figure 8. Class memberships at different truncation levels, 2011.

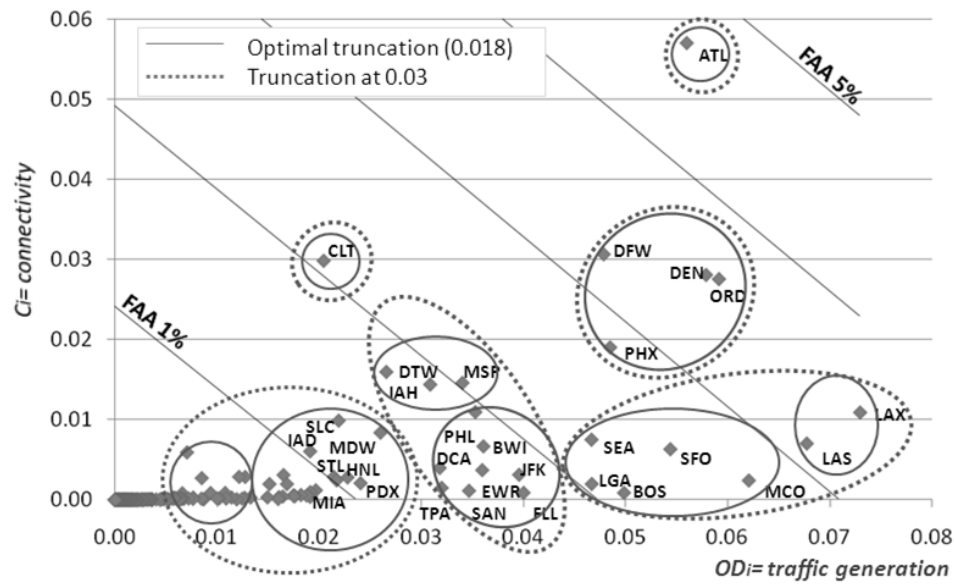


Table 4. Variance decomposition for optimal truncation level (0.0182 dissimilarity).

Variance	Absolute	Percent
Within-class	0.000	4.47%
Between-classes	0.000	95.53%
Total	0.000	100.00%

Table 5. Class memberships and centroids for optimal truncation level.

<i>Class</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	
Objects	1	4	2	5	1	3	8	22	30	
Minimum distance to centroid	0.000	0.005	0.003	0.004	0.000	0.001	0.000	0.001	0.001	
Average distance to centroid	0.000	0.007	0.003	0.006	0.000	0.003	0.004	0.003	0.003	
Maximum distance to centroid	0.000	0.009	0.003	0.010	0.000	0.004	0.007	0.009	0.005	
Class members	ATL	ORD DEN DFW PHX	LAX LAS	MCO SFO SEA BOS LGA	CLT	MSP DTW IAH	PHL JFK BWI FLL EWR SAN DCA TPA	MDW SLC PDX HNL RDU IAD MIA MSY MKE STL MCI SAT OAK PIT HOU RSW SNA	SMF AUS BNA SJU CLE OGG BUR GEG PVD MHT OMA ELP BHM BOI TUL	DAL ONT RIC LGB GEG PVD MHT OMA ELP BHM BOI TUL
<i>Centroid</i>	<i>ATL</i>	<i>DEN</i>	<i>LAS</i>	<i>SFO</i>	<i>CLT</i>	<i>DTW</i>	<i>EWR</i>	<i>OAK</i>	<i>OMA</i>	
OD- traffic generation	0.056	0.058	0.068	0.054	0.020	0.031	0.036	0.019	0.008	
C- connectivity	0.057	0.028	0.007	0.006	0.030	0.014	0.004	0.001	0.000	

In addition to the hubs, this alternative classification has three additional groups for "traffic generators" (Table 7). In the first tier, we find the main airports serving the largest metropolitan areas in the US, for which a representative airport would be San Francisco (SFO). In the second tier are found airports such as Baltimore-Washington or Newark. The remaining airports are grouped in the third tier.

Table 6. Clusters criteria and representative airports.

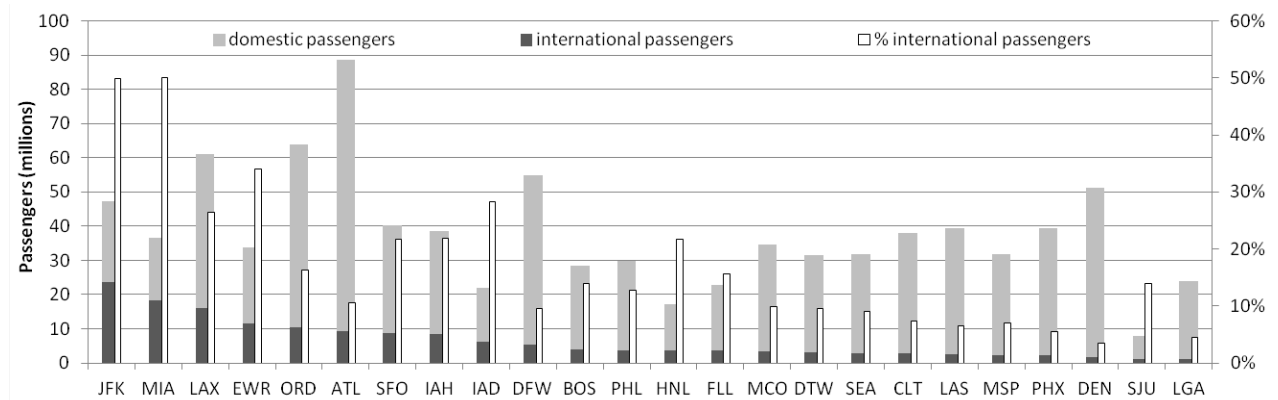
<i>Hubs</i>	<i>Representative</i>	<i>OD%</i>	<i>C%</i>
1 st tier	Atlanta	>5%	>5%
2 nd tier	Denver	>5%	>2%
3 rd tier	Charlotte	>2%	>2%
<i>Traffic generators</i>	<i>Representative</i>	<i>OD%</i>	<i>C%</i>
1 st tier	San Francisco	>5%	-
2 nd tier	Baltimore-Washington	>3%	-
3 rd tier	Oakland	>1%	-

Hence, Table 6 summarizes the alternative classification for regulatory purposes. The values are based on the edges of the cluster described above. It is worth highlighting the simplicity and similarity with the current FAA method, the availability of the data to perform the calculations, and its ready applicability.

Nevertheless, it is important to acknowledge a limitation that raises from the dataset. Note the odd location of large international gateways such as New York-JFK, Miami (MIA) or Washington-Dulles (IAD), which show low levels of connectivity. It seems difficult to justify that these important airports are classified as second or third tier traffic generators. Clearly, this is related to the absence of international markets in the BTS dataset. As a result, all these large gateways are characterized here only by their contribution to domestic markets. We believe that this issue could be overcome by using supply data followed by correction algorithms, yet this remains out of the scope of this paper and does not invalidate its main contributions, which are in relation to the flow centrality measure and the clustering analysis. In addition, gateways are easily identifiable by their substantial amount of

international passengers and their dominant position within the network of international connections (Figure 9). They tend to be located in large urban regions and have a more stable traffic since they often have emerged at the convergence on inland transport systems (Rodrigue et al., 2006), while hubs can disappear if the carrier withdraws the services. Hence, since gateways can be singled out and the average percentage of international passengers at US airport is 2%, the effects of international transfers for non gateway airports can be neglected (Adikariwattage et al., 2012).

Figure 9. Largest international gateways in the US. Source: Own elaboration from the Bureau of Transport Statistics.



5. Conclusions

In summary, this paper develops a flow-based indicator of airport connectivity and measures its sensibility to airline de-hubbing. We argue that, under perfect information, flow centrality should collapse into a simple ratio between connecting passengers and total network passengers that do not originate or terminate the base airport. An application to the US domestic network has been provided, using demand data to perform a detailed time-series analysis of airport connectivity patterns between 1993 and 2012. The flow-based indicator is then used to define an alternative airport classification method within the context of the Federal Aviation Administration's National Plan of Integrated Airport Systems (NPIAS).

For the sensibility analysis, several de-hubbing cases are examined and in which the flow-based indicator is shown to be much more sensitive than other indicators that have been used in the same context such as degree centrality and betweenness centrality. This is related to the fact that these topological measures only take into account the number of established traffic links without considering the density of traffic flows. Thus, we conclude that flow-based centrality could be used as the standard demand-based indicator to measure actual airport connectivity.

From the policy perspective, the suitability of this indicator to serve as a criterion for airport classification in the US domestic network was discussed. The major requirement for the regulator would be to set the thresholds that define the airport categories, which can be easily obtained using data clustering techniques, such as the we have used.

From a methodological point of view, further research could try to investigate ways to cover the limitations on the availability of international demand data. This might be overcome by using supply data followed by correction algorithms.

From an analysis point of view, further research could focus on applying the flow-based indicator to do much in-depth demand-based analysis of airline de-hubbing cases and, in particular, on the variables that have an impact on airport recovery. Also, with regard to the airport clustering methods, there is scope for more studies looking into the usefulness of this method for the definition of policies and regulatory norms, as well as airport performance evaluation.

References

Adikariwattage, V., de Barros, A.G., Wirasinghe, S.S., Ruwanpura, J., 2012. Airport classification criteria based on passenger characteristics and terminal size. *Journal of Air Transport Management* 24, 36-41.

- Bagler, G., 2008. Complex network view on performance and risks on airport networks. arXiv:0805.0924v1, 7 May 2008. Available on-line at http://arxiv.org/PS_cache/arxiv/pdf/0805/0805.0924v1.pdf.
- Berger, A., Müller-Hannemann, M., Rechner, S., Zock, A., 2011. Efficient computation of time-dependent centralities in air transportation networks. *Lecture Notes in Computer Science* 6552, 77-88.
- Bhadra, D., 2009. Race to the bottom or swimming upstream: performance analysis of US airlines. *Journal of Air Transport Management* 15, 227-235.
- Bootsma, P. D., 1997. Airline flight schedule development; analysis and design tools for European hinterland hubs. PhD thesis, University of Twente, Utrecht.
- Budde, A., de Wit, J., Burghouwt, G., 2008. Borrowing from behavioural science: a novel method for the analysis of indirect temporal connectivity at airport hubs. *Proceedings of the Air Transport Research Society Conference*, Athens, 6–10 July 2008.
- Burghouwt, G., Hakfoort, J., 2001. The evolution of the European aviation network, 1990–1998. *Journal of Air Transport Management* 7, 311–318.
- Burghouwt, G., 2007. Airline network development in Europe and its implications for airport planning. Aldershot, Ashgate.
- Burghouwt, G., Redondi, R., 2013. Connectivity in Air Transport Networks. An Assessment of Models and Applications. *Journal of Transport Economics and Policy* 47, 35-53.
- Button, K., 2002. Debunking some common myths about airport hubs. *Journal of Air Transport Management* 8, 177-188.
- Calinski, T., Harabasz, J. 1974. A dendrite method for cluster analysis. *Communications in Statistics* 3, 1-27.
- Caves, R.E., 1997. European airline networks and their implications for airport planning. *Transport Reviews* 17(2), 121–144.

- Cronrath, E., Arndt, A., and Zoch, A., 2008. Does size matter? The importance of airports in the European and German air transport network. Proceedings of the Air Transport Research Society Conference, Athens, 6–10 July 2008.
- Danesi, A., 2006. Measuring airline hub timetable co-ordination and connectivity: definition of a new index and application to a sample of European hubs. *European Transport* 34, 54–74.
- Dennis, N. P., 1994a. Scheduling strategies for airline hub operations. *Journal of Air Transport Management*, 1(2), 131–44.
- Dennis, N. P., 1994b. Airline hub operations in Europe. *Journal of Transport Geography*, 2(4), 219–33.
- Doganis, R., 2010. Flying off course: the economics of international airlines (4th Ed.). Routledge. London.
- Doganis, R. and N. Dennis, 1989. Lessons in hubbing. *Airline Business*, March 1989, 42–7.
- Everitt, B., Landau, S., and Leese, M., 2001. Cluster Analysis 4th ed. Wiley, London.
- FAA, 2011. Federal Aviation Administration National Plan of Integrated Airport System (NPIAS) 2011-2015 Report. [online] http://www.faa.gov/airports/planning_capacity/npias/reports/
- Freeman, L., 1977. A set of measures of centrality based on betweenness. *Sociometry* 40, 35-41.
- Freeman, L., 1978. Centrality in social networks: Conceptual clarification. *Social Networks* 1, 215-239.
- Freeman, L., Borgatti, S., and White, R., 1991. Centrality in valued graphs: a measure of betweenness based on network flow. *Social Networks* 13, 141-154.
- Galle, K., Ale, J., Hossain, M., Moliterno, M., Rowell, M., Revenko, N., Rogerson, E., Tucker, S., Crowther, K., Lambert, J., and Haimes, Y., 2010. Risk-based airport selection

- for runway safety assessments through the development and application of systems-driven prioritization methodologies. Proceedings of the 2010 IEEE Systems and Information Engineering Design Symposium, University of Virginia.
- Guida, M., and Maria, F., 2007. Topology of the Italian airport network: A scale-free small-world network with a fractal structure?. *Chaos, Solitons and Fractals* 31, 527–536.
- Guimerà , R., Mossa, S., Turtschi, A., and Amaral, L. A., 2005. The worldwide air transport network: anomalous centrality, community structure and cities' global roles'. *Proceedings of the National Academy of Sciences* 102(22), 7794–9.
- Jessop, A., 2012. A decision aid for finding performance groups. *Benchmarking: An International Journal* 19 (3), 325-339
- Jia, T., Jiang, B., 2012. Building and analyzing the US airport network based on en-route location information. *Physica A* 391, 4031–4042.
- Liu, Z-J., Debbage, K., Blackburn, B., 2006. Location determinants of major US air passenger markets by metropolitan area. *Journal of Air Transport Management* 12, 331-341.
- Madas, M., and Zografos, K., 2008. Airport capacity vs. demand: Mismatch or mismanagement? *Transportation Research A* 42, 203–226.
- Malighetti, P., Paleari, S., and Redondi, R., 2008. Connectivity of the European airport network: 'Self-help hubbing' and business implications. *Journal of Air Transport Management* 14(2), 53–65.
- Malighetti, P., Paleari, S., Redondi, R., 2009. Airport classification and functionality within the European network. *Problems and Perspectives in Management* 7, 183-196.
- Matsumoto, H., Veldhuis, J., de Wit, J., and Burghouwt, G., 2008. Network performance, hub connectivity potential, and competitive position of primary airports in Asia/Pacific region. *Journal of Airport Management* 3, 384-400.

- Nieminem, J., 1974. On centrality in a graph. *Scandinavian Journal of Psychology* 15, 322-336.
- Oum, T.H., Zhang, A., Zhang, Y., 1995. Airline network rivalry. *Journal of Economics* 18, 836-857.
- Paleari, S., Redondi, R., and Malighetti, P., 2010. A comparative study of airport connectivity in China, Europe and US: which network provides the best service to passengers?. *Transportation Research E* 46(2), 198–210.
- Redondi, R., Malighetti, P., Paleari, S., 2012. De-hubbing of airports and their recovery patterns. *Journal of Air Transport Management* 18, 1-4.
- RITA, 2013. Airline Origin and Destination Survey DB1B. Research and Innovative Technology Administration. US Bureau of Transportation Statistics, Washington DC. URL: http://www.rita.dot.gov/bts/data_and_statistics/index.html
- Reggiani, A., Nijkamp, P., Cento, A., 2008. Connectivity and competition in airline networks: a study of Lufthansa's network. In Vervest, P. H. M., Zheng, L., and van Liere, D. W., (eds), *The Network Experience*, Springer, Berlin, 141–63.
- Rodrigue, J-P., Comtois, C., Slack, B., 2006. *The Geography of Transport Systems*. Oxton, Routledge. Second Edition.
- Rodríguez-Déniz, H., 2012. Using SAS® to Measure Airport Connectivity: An Application of Weighted Betweenness Centrality for the FAA National Plan of Integrated Airport Systems (NPIAS). *Proceedings of the SAS Global Forum 2012*, Paper 162-2012.
- Rodríguez-Déniz, H., Voltes-Dorta, A. 2014. A frontier-based hierarchical clustering for airport efficiency benchmarking. *Benchmarking: an international journal*. Article in Press.
- Sarkis, J., and Talluri, S., 2004. Performance based clustering for benchmarking of US airports. *Transportation Research A* 38, 329–346.

- Suau-Sanchez, P., and Burghouwt, G., 2012. Connectivity levels and the competitive position of Spanish airports and Iberia's network rationalization strategy, 2001-2007. *Journal of Air Transport Management* 18, 47-53.
- Veldhuis, J., 1997. The competitive position of airline networks. *Journal of Air Transport Management* 3(4), 181-8.
- Wang, J., Mo, H., Wang, F., Jin., F., 2011. Exploring the network structure and nodal centrality of China's air transport network: a complex network approach. *Journal of Transport Geography* 19, 712-721.
- Xu, Z., Harriss, R., 2008. Exploring the structure of the U.S. intercity passenger air transportation network: a weighted complex network approach. *Geo Journal* 73, 87-102.
- Xu, R, Wunsch, D., 2005. Survey of Clustering Algorithms. *IEEE Transactions on Neural Networks* 16 (3), 645-678.
- Zeng, X., Tang, X., and Jiang, K., 2011. Empirical study of chinese airline network structure based on complex network theory. *Journal of Transportation Systems Engineering and Information Technology* 11(6), 175-181.

APPENDIX A.

Figure B.1 Full airport dendrogram (optimal truncation level: 0.0182 dissimilarity).

